

CRF Errors Corrected by the STIC Systems Branch

CRF Processing Date: 7-11-02

Edited by: M SPENCER

Verified by: _____

(STIC staff)

41 70590
6/2003

Serial Number: 10/042,665A

 Changed a file from non-ASCII to ASCII Changed the margins in cases where the sequence text was "wrapped" down to the next line. Edited a format error in the Current Application Data section, specifically: Edited the Current Application Data section with the actual current number. The number inputted by the applicant was the prior application data; or other _____. Added the mandatory heading and subheadings for "Current Application Data". Edited the "Number of Sequences" field. The applicant spelled out a number instead of using an integer. Changed the spelling of a mandatory field (the headings or subheadings), specifically: Corrected the SEQ ID NO when obviously incorrect. The sequence numbers that were edited were: Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited: Corrected subheading placement. All responses must be on the same line as each subheading. If the applicant placed a response below the subheading, this was moved to its appropriate place. Inserted colons after headings/subheadings. Headings edited included: Deleted extra, invalid, headings used by an applicant, specifically: Deleted: non-ASCII "garbage" at the beginning/end of files; secretary initials/filename at end of file; page numbers throughout text; other invalid text, such as _____. Inserted mandatory headings, specifically: Corrected an obvious error in the response, specifically: Edited identifiers where upper case is used but lower case is required, or vice versa. Corrected an error in the Number of Sequences field, specifically: A "Hard Page Break" code was inserted by the applicant. All occurrences had to be deleted. Deleted ending stop codon in amino acid sequences and adjusted the "(A)Length:" field accordingly (error due to a PatentIn bug). Sequences corrected: Other:

*Examiner: The above corrections must be communicated to the applicant in the first Office Action. DO NOT send a copy of this form.

3/1/95



OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002

TIME: 09:11:59

Input Set : A:\ptoms.txt
 Output Set: N:\CRF3\07112002\J042665A.raw

3 <110> APPLICANT: Schupp, Thomas
 4 Toupet, Christine
 5 Engel, Nathalie
 7 <120> TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
 9 <130> FILE REFERENCE: 4-21001/B/C1
 11 <140> CURRENT APPLICATION NUMBER: 10/042,665A
 12 <141> CURRENT FILING DATE: 2002-01-09
 14 <150> PRIOR APPLICATION NUMBER: 09/242,744
 15 <151> PRIOR FILING DATE: 1999-03-24
 17 <150> PRIOR APPLICATION NUMBER: PCT/EP97/04495
 18 <151> PRIOR FILING DATE: 1997-08-18
 20 <160> NUMBER OF SEQ ID NOS: 9
 22 <170> SOFTWARE: PatentIn Ver. 2.1
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 5676
 26 <212> TYPE: DNA
 27 <213> ORGANISM: Amcylatopsis mediterranei
 29 <400> SEQUENCE: 1
 30 ggtacccgggt gttcgcgacg ggttcgacg aggcttcgaa gcagctggac gtctgtctgg 60
 31 ccggccgtgc cgggcaccgc gtgcgggacg tcgtgtctgg cgaagtgcgc gccaacccg 120
 32 ggctgctgaa ccagacggtc ttccaccaag cccggctgtt cgcgggtggag agcgcgtgt 180
 33 tccggctcgc cgaatccctgg ggtgtccggc cggacgtggt gtcggccac tccatcgggg 240
 34 agatcaccgc cgcgtatgcc gccccgcgtct ttcgtgtgcc ggacgcgcgc cggatcgtcg 300
 35 cggcgcgcgg cccgctgtatg caggcgctgg cggccggcgg ggcgatggtc gccgtcgccg 360
 36 cctccgaagc cgagggtggcc gaactgtctg gcaacggcgt ggaactgcgc gccgtcaacg 420
 37 gcccattcggc ggttagtcctt tccggggacg cggacgcggc ctgcgcggc gcccggcga 480
 38 tgcgcgagcg cgggcacaag accaagcagc tcaagggttc gcacgcgttc cactccgcgc 540
 39 ggatggcgcc gatgttgcgcg gagttcgcgcg ccgagctggc cggcgtgacg tggcgcgagc 600
 40 cggagatccc ggtggcttc aacgtgaccc gcccgttcgc cgagccgcgc gaactgaccg 660
 41 agccgggcta ctggggccgag cacgtgcggc ggcgggtgcg ttgcgcggag ggcgtcgccg 720
 42 cccgcacggc gtccggcgcc tgcgtgttcg tggagctcgg gccggggccg ggcgtgaccg 780
 43 ccctcgtaa ggagacggcc gaggtcacct gcgtgcggc cctgcgggac gaccggccgg 840
 44 aggtcaccgc gctgatcacc ggggtcgcgc agctgttcgt ccgcgggggtt gcggtcgatt 900
 45 ggcggccct gctgcgcgcg tcaccgggt tcgtgcaccc gccaaggtac gccttcgacc 960
 46 agcagcacta ttggctgcag cccgcgcgc aggccacggc cgcggccctcg ctcggcagg 1020
 47 tcgcggccga ccacccgtg ctgggcgcgg tggccggct gcccgcgtcg gacggctgg 1080
 48 tcttcaccc tcggctgtca ttgaaatcgc acccggtgc ggcgcaccac gtcatcgccg 1140
 49 gggtggtgct cgtgcgcggc accgggtcgc tcgagctggc cgtccggcc ggggacgagg 1200
 50 ccggctgcgc ggtccctgaa gaactcgtca tcgaggctcc gtcggtcgtc cccgaccacg 1260
 51 gcgggggtccg gatccaggc gtcgtgggg caccggggga gaccgggtcg cgcgcggtcg 1320
 52 aggtgtactc cctgcgcgcg gacgcgcgtg ccgaagtgtg ggcggccac gccaccgggt 1380
 53 tcctggctgc gacgcgcgtc cagcacaagc cgttcgactt caccgcctgg cgcgcgcggc 1440
 54 gcgtcgagcg cgtcgacgtc gaggactt acgacggctt cgtcgaccgc gggtaacgcct 1500

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002
TIME: 09:11:59

Input Set : A:\ptoms.txt
Output Set: N:\CRF3\07112002\J042665A.raw

```

55 acgggccgtc gttccgggac ctgcggcggt tggccggcg cggcgtacgaa gtgttcgccc 1560
56 aggtcgccct ggccgaggac gaccgcgcgg acgcggcccc gttcggcatc caccggcc 1620
57 tgctggacgc cgccctgcac gcgggcattgg ccgggtccac caccacggaa gagcccgccc 1680
58 ggccgggtgct gccgttgcgc tgaacggcc tgggtctgca cggccggccgg gcttccggc 1740
59 tgcgggtccg gtcgcggcc acgggtccgg acgcgcgttc gtcggaggcc gccgacgagg 1800
60 ccggcggtct cgttgtgacg gcggactcgc tgggtctccgg gccgggtctg gccaacacg 1860
61 tgggcgcggc ggcgaaccac gacgcgttg tccgcgttga gtggaccgag atttctcg 1920
62 ctggagacgt tccggccggac cacgtcgaag tgcgtcgaagc cgtccggcggag gatccctgg 1980
63 aactgaccgg ccgggtctg gaggccgtgc agacctggct cgccgacgca gccgacgacg 2040
64 ctcgccttgt cgttgtgacc cggccggccg tccacgaggt gactgaccgg gccgggtccg 2100
65 cgggtgtgggg cctgatccgg gccgcgcagg cggaaaaccc ggaccggatc gtgtctgg 2160
66 acaccgacgg tgaagtggcc ctaggccggg tgctggccac cggcgtacggcc caaacagccg 2220
67 tccgaggcgc cacgtgttc gccccggcgc tggccggcgc cgaggccggc gaggcacccg 2280
68 cagtgaccgg cgggacggtc ctgtatctgg gcgcgggttc gtcggccggc ctcacccggcc 2340
69 ggcaccttgtt cggccggcgc gtagtccggc ggctgggtgt cgtcggccgc cgtggccccc 2400
70 acgcccacgg catggccgaa ctgaccgtt aactcatcgc tcaggccgcg gaggtcgccg 2460
71 tagtcgtttt cgacctggcc gaccgggacc aggtccgggt actgtctggcc gagcaccggc 2520
72 cgaacgcgtt cgtgcacacg gccgggtttc tcgacgacgg cgtcttcgag tcgctgacgc 2580
73 gggagccggct ggccaagggtc ttgcggccca aagttactgc tgccaatcac ctcgacgagc 2640
74 tgacccgcga actggatctt cgcgcgttc tggtttctc ctccgcctcc ggggtcttcg 2700
75 gctccggccgg gcagggcaac tacgcccgtt ccaacgccta cttggacgccc gtggtcgcca 2760
76 accggccgggc cgcggccctt cccggcacat cgctggctt gggctgttgg gaacagaccg 2820
77 acgggatgac cgcgcacccgc ggcgacggcc accaggccgcg ggcgagtcgc ggcgggtcc 2880
78 tcgcgcatttc acccgccgaa ggcattggacg tggccgttc acgcggccgc gggctctcg 2940
79 tcccggtcaa gctggacctt cgcacggacc ggcggccggc gacgggtccg cacctgtgc 3000
80 gcggccttgtt cggccggggc cggcagcagg cccgtccggc gtccactgtt gacaacggac 3060
81 tggccggccgg actcgccggg ctgcggccgg cggagcagga ggcgctgtc ctcgacgtcg 3120
82 tcgcacgcgca ggtcgccgtt tgctcgccggc acgcggccgc ggaggccgtc cgcgcggaca 3180
83 cggcggtcaa ggacaccggc ttgcacttcgc tgacgtcggt ggaactgcgc aaccggctgc 3240
84 gcgaggccgg cgggctgttgg ctgcggccgcg cgctcgctt cgactaccgg acgcgggtcg 3300
85 cgctggcccg ctacctgtgt gacgaattcg gcgcacacggt ggcaacaact cgggtggcca 3360
86 cccggccgcg acggacgcgc ggcgagccgc tcgcacgtt cggcatggcc tgccggctgc 3420
87 cggcgccgggtt caccgttccgaaaggccgtt ggcgccttgg ggcgcacggc ctcgaaggcc 3480
88 tgttccctt ccccgaggac cggggcttggg acctggagaa ctgttgcac gacgacccgg 3540
89 accgctccgg cacgacgtac accagccggg gcccgttctt cgacggccgc ggcctgttgc 3600
90 acgcggccgtt ttccggatt tcgcggccgcg aggctgttgc catggaccgg cagaacggc 3660
91 tgctgtcgat ggcggcttgg gaagccctcg aaggcaccgg tgctgaccgg ggctctgttga 3720
92 agggccgcga cgtcggttgg ttcggccggg tgcgttggat gggctatggg atggccggcg 3780
93 atccggccga actggccggg ttcgcgttgc cggccggccgc ttgcgtcgatc gtctcgcc 3840
94 gagtctcgat cgtcttcggg ttgcgttgc cggccgttgc gatcgacacgg gtttgcgttgc 3900
95 cgtcgcttgtt ggcgtatgcac ctggccggggc aggctgttgcg gacggccggag tgctcgatgg 3960
96 ccctggccgg tggcgatccgcg gtcgttggat gggccggccac gtcgttggat ttgcgttgc 4020
97 agcgcggccct ggccggccgc ggcggccgttgc aggcttgcgc cgaaggccgc gacggccacgg 4080
98 gctggccggc gggcgccgttgc gtcgttgc tggagccggc gtcgttggcc cgcgcacggc 4140
99 ggcacccgggt gctggccgttgc ctgcgttgc ggcgttgc gacggccggat ggcgttccacgg 4200
100 ggcgttgcgc ccccaacccggg cgcgttgcacg aacgggttgc ggcggccggcc ctggccggcg 4260
101 cccggccgttgc accgttccgtt gtcgttgc tggagccggc ggcgttccgg acggccgttgc 4320
102 ggcacccgttgc gaggccgttgc ggcgttgc ccacccgttgc aaggccgttgc gacggccggat 4380
103 cggccgttgc gctggccgttgc gtcgttgc gtcgttgc acggccgttgc cgcgttccacgg 4440

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002
TIME: 09:11:59

Input Set : A:\ptoms.txt
Output Set: N:\CRF3\07112002\J042665A.raw

104 tggccgggtt gatcaagatg gtgcaggcgc tgccacgg cgtcatgccg cccaccctgc 4500
 105 acgtggaccg gcccaccagc caggtcgact ggtccgggg ggccgtcgaa gtgtgaccg 4560
 106 aggcacggga gtggcccgaa aacggccgtc cgccggccggc cgggggtgtcc tcgttcggga 4620
 107 tcagcggcac gaacgcccac ctgatcatcg aagaagcacc ggccgagcca cagcttgcgg 4680
 108 gaccaccgccc ggacggcggt gtggcccgcc tggtcgctc ggctcgccagc cccgggtgccc 4740
 109 tggccggtaa ggcggtcggt ctggccacgt tcctcgccga cggggccctt tccgacgtcg 4800
 110 cccgtgcgtc gacgagccgc gccctgttcg gcgagcgcgc ggtcgctgtc gcgattcgg 4860
 111 ccggaggaagc ccggccgggt ctggggcgaac tggcccgccgg cgaagacgcg ccgggcctgg 4920
 112 tcccgccgccc ggtgcccccg tccggccctgc cgggcaagct cgtgtgggtg ttccccgggc 4980
 113 aggggacgca gtgggtgggc atggggcccg aactcctcga agagtctccg gtgttcggcc 5040
 114 agcggatcgc cgagtggtcgcc gccgcgtcg agccgtggat cggctggtcg ctgttcgacg 5100
 115 tcctccgtgg cgacgggtgac ctgcgtcggt tcgatgtgtc gcagcccgcg tgctttgcgg 5160
 116 tgcgtgtcggtt cttggcccggt gtgtggtcct cggccgggggt ggtccccat gcggtgtcg 5220
 117 gccactccca gggtgagatc gccgcggcgt gcgtgtcggt tgcttgcgtc ctggaggatg 5280
 118 cggcgaagggt ggttgcctg cgcagccagg ccatcgccgc gaagctctcc ggcgcggcgc 5340
 119 ggatggcttc ggtcgccctt ggcgaagccg atgtgggtgc gcggtggcg gacgggggtcg 5400
 120 aggtggctgc cgtcaacggt cccggcgtccg tggtgatcgc gggggatgcc cagggccctcg 5460
 121 acgaaacgcgt ggaagcgttgc tccgggtcggtt gaatccgggc tcggcgggtg gcggtggact 5520
 122 acgcctcgttgc caccoggcgttgc gtcgaagaca tcgaagacac ctcgcggcaa gcgtggccg 5580
 123 ggatcgacgc ccggccggccgttgc tggtgccttgc tcctctccac ctcaccggc gagtggatcc 5640
 124 gggacgaggg cgtcgtggac ggcggctact ggtacc 5676
 127 <210> SEQ ID NO: 2
 128 <211> LENGTH: 1875
 129 <212> TYPE: PRT
 130 <213> ORGANISM: Amycolatopsis mediterranei
 132 <400> SEQUENCE: 2
 133 Tyr Pro Val Phe Ala Thr Ala Phe Asp Glu Ala Cys Glu Gln Leu Asp
 134 1 5 10 15
 136 Val Cys Leu Ala Gly Arg Ala Gly His Arg Val Arg Asp Val Val Leu
 137 20 25 30
 139 Gly Glu Val Pro Ala Glu Thr Gly Leu Leu Asn Gln Thr Val Phe Thr
 140 35 40 45
 142 Gln Ala Gly Leu Phe Ala Val Glu Ser Ala Leu Phe Arg Leu Ala Glu
 143 50 55 60
 145 Ser Trp Gly Val Arg Pro Asp Val Val Leu Gly His Ser Ile Gly Glu
 146 65 70 75 80
 148 Ile Thr Ala Ala Tyr Ala Ala Gly Val Phe Ser Leu Pro Asp Ala Ala
 149 85 90 95
 151 Arg Ile Val Ala Ala Arg Gly Arg Leu Met Gln Ala Leu Ala Pro Gly
 152 100 105 110
 154 Gly Ala Met Val Ala Val Ala Ala Ser Glu Ala Glu Val Ala Glu Leu
 155 115 120 125
 157 Leu Gly Asp Gly Val Glu Leu Ala Ala Val Asn Gly Pro Ser Ala Val
 158 130 135 140
 160 Val Leu Ser Gly Asp Ala Asp Ala Val Val Ala Ala Ala Ala Arg Met
 161 145 150 155 160
 163 Arg Glu Arg Gly His Lys Thr Lys Gln Leu Lys Val Ser His Ala Phe
 164 165 170 175
 166 His Ser Ala Arg Met Ala Pro Met Leu Ala Glu Phe Ala Ala Glu Leu

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002
TIME: 09:11:59

Input Set : A:\ptoms.txt
Output Set: N:\CRF3\07112002\J042665A.raw

167	180	185	190
169 Ala	Gly Val Thr Trp Arg Glu Pro Glu Ile Pro Val Val Ser Asn Val		
170	195	200	205
172 Thr Gly Arg Phe Ala Glu Pro Gly Glu Leu Thr Glu Pro Gly Tyr Trp			
173	210	215	220
175 Ala Glu His Val Arg Arg Pro Val Arg Phe Ala Glu Gly Val Ala Ala			
176	225	230	235
178 Ala Thr Glu Ser Gly Gly Ser Leu Phe Val Glu Leu Gly Pro Gly Ala			
179	245	250	255
181 Ala Leu Thr Ala Leu Val Glu Glu Thr Ala Glu Val Thr Cys Val Ala			
182	260	265	270
184 Ala Leu Arg Asp Asp Arg Pro Glu Val Thr Ala Leu Ile Thr Ala Val			
185	275	280	285
187 Ala Glu Leu Phe Val Arg Gly Val Ala Val Asp Trp Pro Ala Leu Leu			
188	290	295	300
190 Pro Pro Val Thr Gly Phe Val Asp Leu Pro Lys Tyr Ala Phe Asp Gln			
191	305	310	315
193 Gln His Tyr Trp Leu Gln Pro Ala Ala Gln Ala Thr Asp Ala Ala Ser			
194	325	330	335
196 Leu Gly Gln Val Ala Ala Asp His Pro Leu Leu Gly Ala Val Val Arg			
197	340	345	350
199 Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Lys			
200	355	360	365
202 Ser His Pro Trp Leu Ala Asp His Val Ile Gly Gly Val Val Leu Val			
203	370	375	380
205 Ala Gly Thr Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala			
206	385	390	395
208 Gly Cys Pro Val Leu Glu Glu Leu Val Ile Glu Ala Pro Leu Val Val			
209	405	410	415
211 Pro Asp His Gly Gly Val Arg Ile Gln Val Val Val Gly Ala Pro Gly			
212	420	425	430
214 Glu Thr Gly Ser Arg Ala Val Glu Val Tyr Ser Leu Arg Glu Asp Ala			
215	435	440	445
217 Gly Ala Glu Val Trp Ala Arg His Ala Thr Gly Phe Leu Ala Ala Thr			
218	450	455	460
220 Pro Ser Gln His Lys Pro Phe Asp Phe Thr Ala Trp Pro Pro Pro Gly			
221	465	470	475
223 Val Glu Arg Val Asp Val Glu Asp Phe Tyr Asp Gly Phe Val Asp Arg			
224	485	490	495
226 Gly Tyr Ala Tyr Gly Pro Ser Phe Arg Gly Leu Arg Ala Val Trp Arg			
227	500	505	510
229 Arg Gly Asp Glu Val Phe Ala Glu Val Ala Leu Ala Glu Asp Asp Arg			
230	515	520	525
232 Ala Asp Ala Ala Arg Phe Gly Ile His Pro Gly Leu Leu Asp Ala Ala			
233	530	535	540
235 Leu His Ala Gly Met Ala Gly Ala Thr Thr Glu Glu Pro Gly Arg			
236	545	550	555
238 Pro Val Leu Pro Phe Ala Trp Asn Gly Leu Val Leu His Ala Ala Gly			
239	565	570	575

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002
TIME: 09:11:59

Input Set : A:\ptoms.txt
Output Set: N:\CRF3\07112002\J042665A.raw

```

241 Ala Ser Ala Leu Arg Val Arg Leu Ala Pro Ser Gly Pro Asp Ala Leu
242      580      585      590
244 Ser Val Glu Ala Ala Asp Glu Ala Gly Gly Leu Val Val Thr Ala Asp
245      595      600      605
247 Ser Leu Val Ser Arg Pro Val Ser Ala Glu Gln Leu Gly Ala Ala Ala
248      610      615      620
250 Asn His Asp Ala Leu Phe Arg Val Glu Trp Thr Glu Ile Ser Ser Ala
251      625      630      635      640
253 Gly Asp Val Pro Ala Asp His Val Glu Val Leu Glu Ala Val Gly Glu
254      645      650      655
256 Asp Pro Leu Glu Leu Thr Gly Arg Val Leu Glu Ala Val Gln Thr Trp
257      660      665      670
259 Leu Ala Asp Ala Ala Asp Asp Ala Arg Leu Val Val Val Thr Arg Gly
260      675      680      685
262 Ala Val His Glu Val Thr Asp Pro Ala Gly Ala Ala Val Trp Gly Leu
263      690      695      700
265 Ile Arg Ala Ala Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu Asp
266      705      710      715      720
268 Thr Asp Gly Glu Val Pro Leu Gly Arg Val Leu Ala Thr Gly Glu Pro
269      725      730      735
271 Gln Thr Ala Val Arg Gly Ala Thr Leu Phe Ala Pro Arg Leu Ala Arg
272      740      745      750
274 Ala Glu Ala Ala Glu Ala Pro Ala Val Thr Gly Gly Thr Val Leu Ile
275      755      760      765
277 Ser Gly Ala Gly Ser Leu Gly Ala Leu Thr Ala Arg His Leu Val Ala
278      770      775      780
280 Arg His Gly Val Arg Arg Leu Val Leu Val Ser Arg Arg Gly Pro Asp
281      785      790      795      800
283 Ala Asp Gly Met Ala Glu Leu Thr Ala Glu Leu Ile Ala Gln Gly Ala
284      805      810      815
286 Glu Val Ala Val Val Ala Cys Asp Leu Ala Asp Arg Asp Gln Val Arg
287      820      825      830
289 Val Leu Leu Ala Glu His Arg Pro Asn Ala Val Val His Thr Ala Gly
290      835      840      845
292 Lys Val Phe Ala Pro Lys Val Thr Ala Ala Asn His Leu Asp Glu Leu
293      850      855      860
295 Thr Arg Glu Leu Asp Leu Arg Ala Phe Val Val Phe Ser Ser Ala Ser
296      865      870      875      880
298 Gly Val Phe Gly Ser Ala Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala
299      885      890      895
301 Tyr Leu Asp Ala Val Val Ala Asn Arg Arg Ala Ala Gly Leu Pro Gly
302      900      905      910
304 Thr Ser Leu Ala Trp Gly Leu Trp Glu Gln Thr Asp Gly Met Thr Ala
305      915      920      925
307 His Leu Gly Asp Ala Asp Gln Ala Arg Ala Ser Arg Gly Gly Val Leu
308      930      935      940
310 Ala Ile Ser Pro Ala Glu Gly Met Glu Leu Phe Asp Ala Ala Pro Asp
311      945      950      955      960
313 Gly Leu Val Val Pro Val Lys Leu Asp Leu Arg Lys Thr Arg Ala Gly

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/042,665A

DATE: 07/11/2002

TIME: 09:12:00

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw